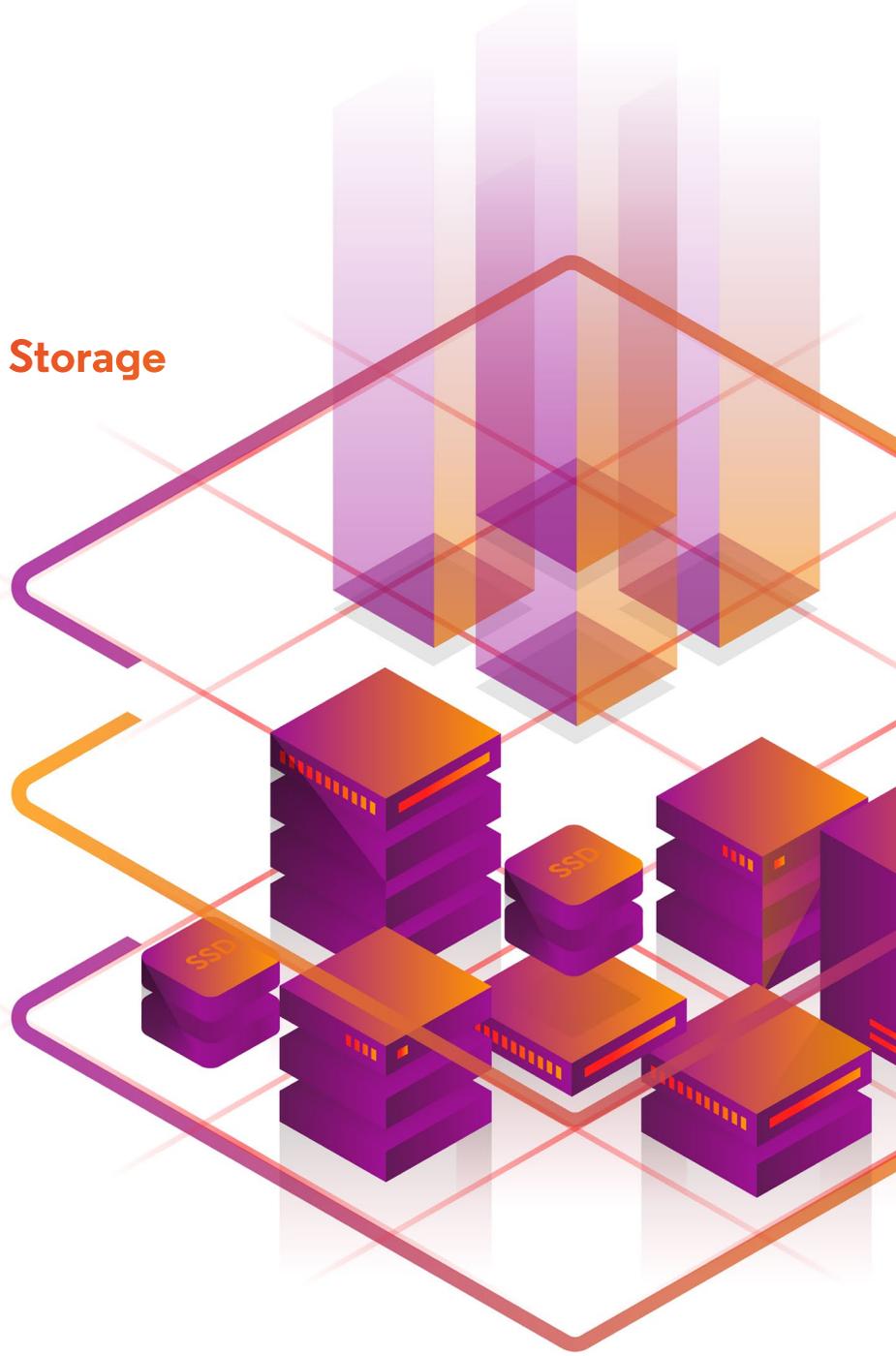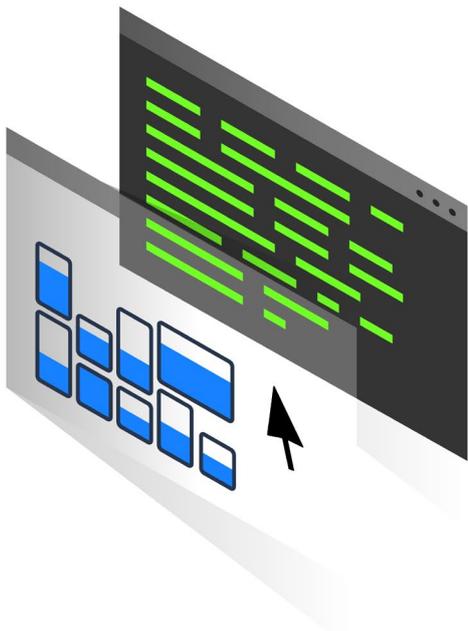# OpenIO

High Performance Object Storage for Big Data, HPC and AI

Data Sheet

Why choose between performance and scalability, when you can have the best of both?

OpenIO is a **software-defined open source object storage** solution ideal for Big Data, HPC and AI. With its **distributed grid architecture** and **self-learning ConsciousGrid™** technology, OpenIO is at the forefront of innovation. Our solution **scales easily without the need for data rebalancing**, and delivers **consistent high performance**, even during scaling operations. OpenIO supports **S3 and Swift** and can be deployed on-premise, cloud-hosted or at the edge, and on **any hardware mix** that you choose.

Storage for the Data-centric World

Here's a question: what do research institutions, financial corporations, healthcare organizations and media conglomerates have in common? The answer is, **they are all built on data**. Whether it be mathematical calculations, medical documentation, multimedia or consumer behavior stats, the data they collect and analyze is at the heart of what they do: it's their most valuable asset.

Now here's the challenge: **the volume of this data is growing**, and there's no end in sight. And **not all companies want to outsource the storage of their most valuable asset. Or get locked** into a storage system where they can't choose the best hardware to scale, or where they would need to pay to repatriate their own data from a public cloud. Plus, to get the most out of their data, **the storage they use must deliver consistent, high-performance throughput**.

WHY S3 OBJECT STORAGE FOR BIG DATA AND HPC?

Scalability: Object storage can scale into hundreds of billions of files. Current HDFS implementations are still limited to hundreds of millions per namenode and do not scale for larger data lakes.

TCO: Object storage costs less to operate than HDFS. Object storage also uses Erasure Coding which is a more efficient and flexible data protection technique.

Performance: In large sequential reads or writes, S3 object storage can be on par with HDFS.
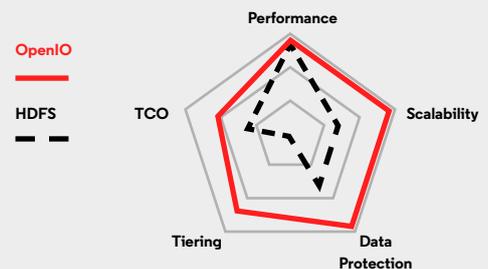
OpenIO Object Storage
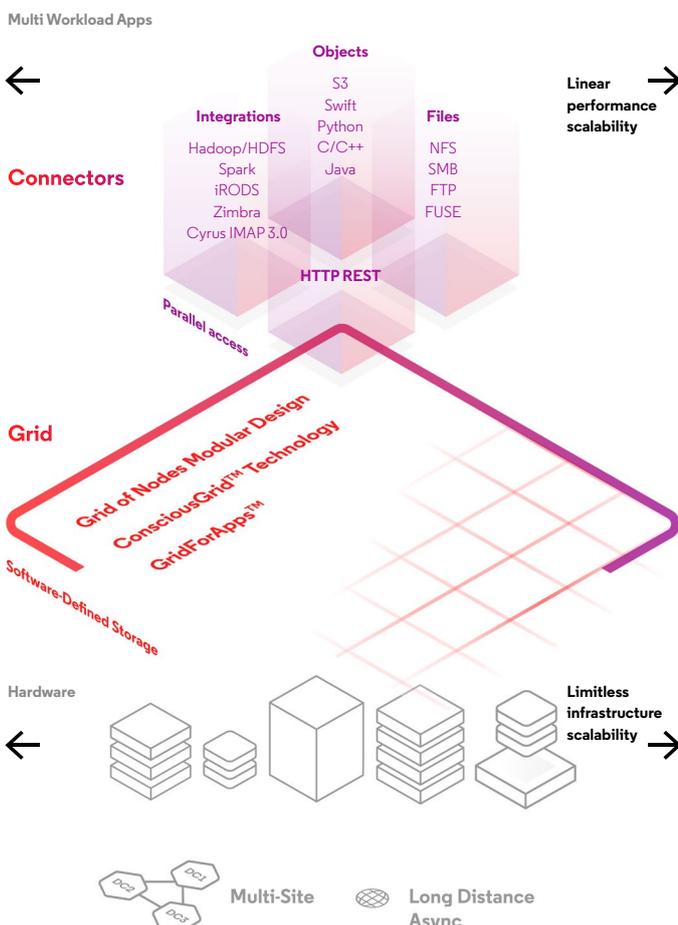
Optimized for Big Data
Integrate seamlessly with your Big Data software stack.

Designed for high performance
Transfer data at 1 Tbps and beyond. Experience consistent high performance, even during scaling operations. Ideal for capacity-intensive and challenging workloads.

Hyper-scalable native object storage
Scale seamlessly from Terabytes to Exabytes. Simply add nodes to expand capacity – without the need for rebalancing data – and increase performance linearly.

Hardware agnostic, lightweight solution
Use servers and storage media that suit your evolving needs. You can combine heterogenous hardware at any time, of different specs, generations, and capacities. Avoid vendor lock-in and achieve real TCO savings.

Technology Highlights

Grid of Nodes Modular Design

The OpenIO cluster layout is organized as a **massively distributed grid of nodes**. This design is more resourceful than a ring-based architecture because it **doesn't require data rebalancing when you change the topology** of a cluster. It is more flexible and more loosely coupled than a ring and it **permits efficient load balancing on heterogeneous nodes**. Using its ConsciousGrid™ technology, OpenIO detects the specific capabilities of each node, so that it can be used at its ideal performance level.

The **data query path is independent of the number of nodes** and you can **add as many as needed**, one by one or in small or large groups, and **without affecting performance or data availability**.
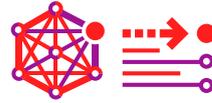
ConsciousGrid™ Technology

This innovative technology uses **real-time metrics from the nodes** (CPU, I/O, capacity…), to automatically discover and place your data on the most appropriate nodes. Since OpenIO discovery and placement method is dynamic, it **avoids the pitfalls of rebalancing** and provides easy decommissioning, so you can scale up and out easily and on any hardware mix that you choose, while **maintaining consistent high performance**.

We integrated a **massively distributed three-level directory** within OpenIO, to facilitate the storage of and access to hundreds of billions of objects. The directory contains the metadata needed to quickly find and **access each object, with the lowest latency**. And in the case of downed servers, objects can be quickly reconstructed so you never lose any data.

**OPENIO INSTANT SCALING:
A BETTER WAY TO SCALE**

- **Add new servers**, one by one or in small or larger groups, as your storage needs grow
- **Nodes are automatically recognized and immediately available**
- **Data does not need to be rebalanced** from old to new nodes
- **Performance remains at a constant high level**, even during scaling operations

---
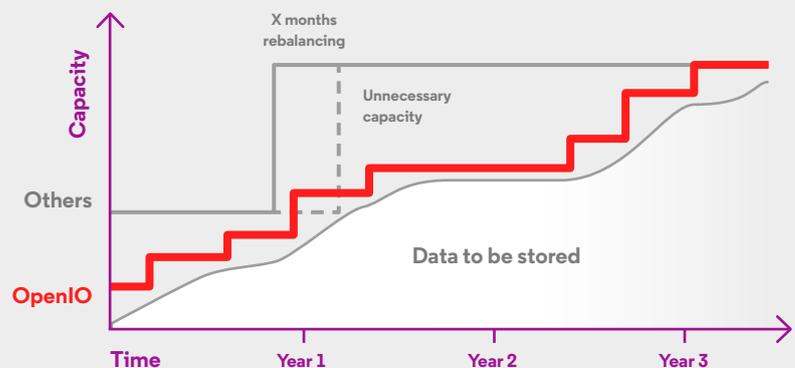
GridForApps™

OpenIO **catches all events** that occur in the cluster and can pass them up in the stack. OpenIO can also perform **real-time event-based metadata indexing and search** directly on the storage infrastructure. More functions like **data augmentation, pattern recognition, machine learning, data filtering and video transcoding** can be activated on customer demand.

Open Source

OpenIO object storage is open core. The heart of the solution and the object APIs are on **GitHub** and available as Ubuntu and CentOS packages. This guarantees the long-term availability of the solution and gives your engineers the opportunity to delve into our technology at any time, to really understand how it works. You can start the conversation with our engineers when you sign up for one of our **subscription plans**, or chat with other users on the **Slack** community.

Features and Specifications

DATA PROTECTION AND INTEGRITY

- **Erasure coding** with multiple **dynamic storage policies** available
- Automatic switch to n-time **replication** for small objects
- **Geo-distributed and stretched** cluster topology
- **Asynchronous long-distance replication** for disaster recovery
- **Encryption of data at rest**
- Data and metadata **integrity checks**, **self-healing** through our integrity loop

DATA MANAGEMENT

- **Storage pools and policies** to **automate** and manage **storage tiering**
- **Automated object versioning**
- On the fly data **compression**, or delayed compression at rest
- **Container snapshots** for backup
- **WORM** mode to create a secure archive (Write Once Read Many, no update or deletion allowed)

SYSTEM MANAGEMENT

- Unified and simplified **CLI** (Command Line Interface)
- **Ansible** deployment tools
- **WebUI** for cluster visualization and routine maintenance operations
- OpenStack **Keystone** for client authentication and user management interface
- **Alerting and monitoring** system, easily expandable through the Prometheus API

OBJECT APIS

- Standard: **Amazon S3, OpenStack Swift**
- Native: **Python, C/C++, Java**

FILE PROTOCOLS

- File Access through OIO-FS: **NFS** with LDAP, **SMB** with Active Directory, local **FUSE**

BIG DATA INTEGRATIONS

- Apache **Hadoop / DistCp** via **S3A** connector, **Spark**, **iRODS**, **HDF5** via Kita

REQUIREMENTS FOR PRODUCTION

- OS: **Linux** (Ubuntu 18.04 LTS, CentOS/RHEL 7+)
- HW agnostic: support for **heterogeneous hardware within clusters**
- Drives: magnetic (**HDDs**) or flash (**SSDs / NVMe**) – flash drives required for metadata
- CPU: **x86** or **ARM**
- RAM: **min 64GB** recommended – depends on number of drives
- Network: **10GbE**

Subscription

Our subscription service is available as **Standard, Standard 24/7 or Premium** plans so you can interact with our experts and receive guidance at any time and when it suits you best.

✓ **24/7 technical support.**

✓ **Continuous assistance for management and fine-tuning** of your production environments.

✓ **End-to-end services** from initial planning to installation, integration, training, and initial tuning.

Find your plan on
www.openio.io/plans